BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

WO 99/64624 PCT/GB99/01691

NUCLEIC ACIDS

The present invention relates to a method of producing a DNA hybrid molecule comprised of a contiguous linear array of three or more sequences of interest (or potential interest) in predetermined relationship to each other.

There are a number of applications in which it would be desirable to provide a DNA hybrid molecule comprising sequences of (potential) interest in predetermined relationship to each other. One such application is in mutation detection where many mutation scanning techniques which provide positional information are capable of scanning up to 1.5kb at a time e.g. CCM, EMC, NIRCA, PTT. However, most genes are organised into short coding sequences (exons) which are dispersed over comparatively large distances. As a result, mutation scanning techniques are not used to their full potential and consequently cost vs benefit considerations have limited their usefulness and favoured less sensitive, size consrained techniques like SSCP and HA. Other alternative methods of circumventing these problems such as RT-PCR are fraught with problems of working with RNA and mutant allele loss in autosomal conditions.

A further application in which it may be desirable to provide a DNA hybrid molecule is DNA sequencing.

Various techniques are already known for producing DNA molecules comprising exons present at diverse locations in the same or different DNA molecule(s). One such technique is Gene "splicing" by overlap extension (SOE) as disclosed in US-A-5 023 171 (see also Horton et al, 1989. Gene, 77, 61-68). In this technique, two gene fragments to be "spliced" together are initially amplified in separate PCR reactions. For each PCR reaction, one of the two primers has a "linker" sequence which is attached to the 5' end of the genomic specific sequence of the

primer and which is designed to be complementary to the linker sequence on a primer of the other PCR reaction.

The amplified gene fragments from the separate PCR reactions are then purified, combined together and subjected to cycles of denaturing, reannealing and strand synthesis. As a result of these cycles, the complementary linker sequences hybridise to each other and are extended to produce a double stranded molecules which comprise the two gene fragments connected by the linker sequences.

The Gene splicing by overlap extension method requires fragments to b joined in a pairwise fashion thus including additional laborious steps of purification and re-amplification for the creation of constructs of more than two fragments. Furthermore, although the Gene Fusion method utilises *Taq* polymerase no account is taken in the primer design of the 3' dA overhangs introduced by *Taq*. Consequently with the primer design employed in US-A-5 023 171, the majority of fragments will have 3' mismatches leading to inefficient fusion of the two fragments.

A further technique for producing DNA hybride molecules is Gene Fusion as disclosed in Nucleic Acids Research, 17, 4895, (1989) (Yon & Freide) where a single linking primer is used to fuse the fragments. This is unlikely to be at all efficient for the fusion of more than two fragments since the PCR products themselves act as substitutes for primers for the whole of the fusion reaction leading to inefficient intermediate formation. Also with this published technique with only two fragments there is contamination by unfused products within the reaction.

A further technique for producing DNA hybrid molecules is disclosed in WO-A-92 15678 (Stratagene) which describes a PCR-based process for generating a library of dicistronic DNA molecules (comprising upstream and downstream cistrons) for producing antibodies. The disclosure contemplates combining, in a single reaction vessel,

(i) a repertoire of first polypeptide genes with a PCR primer pair therefor, one of the primers having a first 5'-terminal non-priming portion providing a linking sequence; and

(ii) a repertoire of second polypeptide genes with a PCR primer pair therefor, one of the primers having a second 5'-terminal non-priming portion.

The first and second 5'-terminal portions are such that they are capable of hybridising to form a duplex encoding a double-stranded cistronic bridge for linking the upstream and downstream cistrons.

Under the conditions of denaturation, reannealing and strand synthesis employed in the reaction, the library of dicistronic molecules is generated.

According to a first aspect of the present invention there is provided a method of producing a hybrid DNA molecule having a sense strand and an anti-sense strand and in which, reading in the 5' to 3' direction, the sense strand has the sequences x_1, x_2.....x_n, where n is greater than or equal to 3, the method comprising the steps of

(i) providing in a single reaction mixture

(a) the sequences x_1, x_2.....x_n and their complementary sequences x_1', x_2'.....x_n', to be assembled into the hybrid molecule,

(b) for each pair of complementary sequences defined in (a) a respective pair of PCR primers each having a priming sequence and which are such that the primers hybridising to the 3' ends of any two sequences (x_i, $x'_{(i+1)}$), where i is 1 to (n-1), have specifically complementary linker sequences

(2) effecting a first stage PCR reaction in which those primers provided with linker sequences are present in limiting concentrations, and

(3) effecting a second stage PCR reaction using a single pair of primers one of which provides the 5'-end of the sense strand and other of which provides the 3'-end of the anti-sense strand of the required hybrid molecule

whereby said hybrid molecule is generated.

In a preferred embodiment of the invention, each of the first and second stage PCR reactions employs *Taq* as a thermostable polymerising enzyme. In this case, those primers having linker sequences are designed such that their linker sequence is connected to their respective priming sequence via an adenosine residue. This takes account of the 3' adenine overhang added by *Taq* at the 3' end of an extended strand. Primers incorporating the extra adenosine residue may of course be used both in conjunction with any other polymerising enzyme which adds a 3'-adenosine overhang at the end of an extended strand and those which do not.

This constitutes an important feature of the present invention according to a second aspect of which there is therefore provided a method of producing a hybrid DNA molecule having a sense strand and an anti-sense strand and in which, reading in the 5' to 3' direction, the sense strand has the sequences $x_1, x_2.....x_n$, where n is greater than or equal to 3, the method comprising the steps of

(i) providing in a single reaction mixture

(a) the sequences $x_1, x_2.....x_n$ and their complementary sequences $x_1', x_2'.....x_n'$, to be assembled into the hybrid molecule,

(b) for each pair of complementary sequences defined in (a) a respective pair of PCR primers each having a priming sequence and which are such that the primers hybridising to the 3' ends of any two sequences (x_i, $x'_{(i+1)}$), where i is 1 to (n-1), have specifically complementary linker sequences connected to their respective priming sequences via an adenine residue, and

(2) effecting a PCR reaction using a thermostable polymerase which may or may not add a 3' adenine overhang to the end of an extended strand.

The invention enables a DNA hybrid molecule comprised of the sequences x_1, x_2.....x_n (and their respective complements x_1', x_2'.....x_n') in predetermined linked relationship to be produced in high yield from the respective individual sequences provided at diverse regions of the same or different DNA molecules. The linker sequences which are used are independent of the sequences which are to be assembled into the intended hybrid molecule. Furthermore the incorporation of the additional adenosine residue for primers to be used in conjunction with *Taq* as the polymerising enzyme avoids 3' mismatches leading to inefficient fusion of fragments.

The sequences x_1, x_2.....x_n may for example be exons which are separated by any distance (possibly of unknown sequence) along a particular DNA molecule whereas in the hybrid molecule produced by the method of the invention the sequences are spaced from each other by a relatively short, and known, sequence of, say, 20 to 30 bases.

Hybrid molecules produced in accordance with the method of the invention may for example be constructed for the purposes of mutation detection. More particularly, the hybrid molecule may comprise a plurality of exons (in which one or more mutations may be present) in a molecule having a length of up to 1.5 kb thus

permitting mutation scanning of all of these exons using a technique such as CCM, EMC, NIRCA or PTT. A further possibility is a hybrid molecule comprised of a plurality of exons which may be constructed to enable more efficient sequencing of exons (as compared to their sequencing the exons at their locations in the endogenous DNA molecule).

A further possibility is that the method of the invention is used for the construction of hybrid genes.

The method of the invention involves two PCR reaction steps which allow the desired DNA hybrid molecule to be produced from the individual sequences x_1, x_2.....x_n (and their complements x_1', x_2'.....x_n') which are provided in a single reaction mixture. It will however be appreciated that, in certain circumstances, only one or other of the sets of complementary the sequences (x_1, x_2.....x_n) or (x_1', x_2'.....x_n') may initially be present at the start of the first PCR reaction, in which case the "missing" complementary sequences are generated *in situ* in the initial PCR reaction stage.

For the first stage of PCR reaction, there is a pair of PCR primers for each set of complementary sequences (x_1, x_1'), (x_2, x_2').....(x_n, x_n') which are to be present in the intended DNA hybrid molecule. The primers hybridising to the 3' ends of the sequences x_1' and x_n may be "standard" PCR primers and may hybridise either to the sequences x_1' or x_n (as the case may be) or to a region external thereto (since the 3' ends of the sequences x_1' and x_n to be incorporated in the hybrid molecule are determined by the primers used in the second stage of the PCR reaction (see below)).

The other primers for the first stage of PCR reaction are in effect modified PCR primers and comprise a priming sequence (i.e. a sequence which will hybridise to the appropriate x or x' sequence in the manner of a standard PCR primer) attached to a linker sequence. (For convenience, primers incorporating linker sequences are

also referred to herein as "linker primers"). The linker sequences are such that the primers which hybridise to the 3' ends of any two sequences (x_i, $x_{(i+1)}$)where i is 1 to (n-1) have linker sequences which are specifically complementary to each other, i.e. these linker sequences will hybridise to each other but not to any other sequences in the reaction mixture. The manner in which the linker sequences function to provide for assembly of the required DNA hybrid molecule will be more fully appreciated from the description given below in relation to the drawing.

Adopting this type of primer design, it is possible to employ linker sequences which are independent of the x and x_i sequences to be incorporated in the desired hybrid molecule so that it is not necessary to ensure that the linker will hybridise to a part of an x or x' sequence to be linked into the hybrid molecule. This design of the linkers is thus not constrained.

The linker sequences may for example be comprised of 20 to 30 bases and are ideally such that they do not have any secondary structure (e.g. "hairpins"). It is preferred that the annealing temperature (Tm) of the complementary pairs are substantially the same and is 2-5°C higher than the annealing temperature of the primers to the x and x' sequences.

Purely by way of example, there is set out below complementary sets of complementary linker sequences which may be incorporated in linker primers for the assembly (using PCR reactions employing *Taq* as the polymerising enzyme) of a hybrid DNA molecule comprised of five sequences x_{1-5} and their respective complements. In the following, the term "specific sequence" is the priming portion of the linker primer.

5'tcatattagccgctgcattgcc-a-specific seq 3'

5'ggcaatgcagcggctaatatga-a-specific seq 3'

5'agccactacccaaactcctgt-a-specific seq 3'

5'acaggagtttgggtagtggct-a-specific seq 3'

5'tgtctcactgaacctgcctacct--a-specific seq 3'

5'aggtaggcaggttcagtgagaca-a-specific seq 3'

5'cctcattaccggctgtcagactg-a-specific seq 3'

5'cagtctgacagccggtaatgagg-a-specific seq 3'

In the above sequences, the "a" represented in bold is an additional adenosine residue incorporated in the primer to take account of the 3' adenosine overhang added by *Taq* at the end of an extended strand.

In the first stage PCR reaction, the linker primers, and preferably also the standard PCR primers for the 3' ends of the sequences x_1' and x_n, are provided in a limiting concentration. A "limiting concentration" is a concentration of primers in a PCR reaction leading to inefficient amplification and such that an increase in concentration results in an increase in product yield. The ratio of primers to single copy template in the limiting PCR reaction may typically be about 1×10^6:1 to about 1×10^8:1 (e.g. about 3×10^7:1). A standard limiting concentration may for example be 40nM. The limiting concentration may however readily be determined empirically for each set of primers used.

Subject to the use of limiting primer concentrations as outlined above, the first stage PCR reaction may be conducted under the conditions (e.g. polymerase enzyme, nucleotides, buffers, temperature cycling etc.) will be understood by those skilled in the art. For example, the temperature cycling may involve stages of denaturation at 90°C, hybridisation at 60°C, and strand synthesis at 72°C.

It is particularly preferred that the polymerase enzyme which is used for the first stage reaction is one formulated so as to be activated by heat (e.g. at 90° to 95°C) so that there are no non-specific hybridisations being extended at low temperature.

Heat activated polymerase enzymes are know to those skilled in the art and examples of such enzymes which may be employed in the method of the invention include AmpliTaq Gold (ex Perkin Elmer) and Platinum *Taq* (ex Gibco BRL).

In a preferred embodiment of the method of the invention, the product of the first stage PCR reaction is treated, prior to the second stage, to prevent the extension of unwanted non-stringent hybridisations between residual primers and the template DNA. This ensures a proper "Hot Start" for the second stage PCR reaction. This treatment may for example comprise an exonuclease digestion involving addition of exonuclease to the first stage reaction mixture followed by incubation of the mixture at 37°C (e.g. for 15 min) and then at 80°C (e.g. for 30 min) so as to remove any single stranded molecules. Alternatively, and more preferably, the first stage reaction mixture is cooled, e.g. at -20°C, to inactivate residual DNA polymerase activity.

In the second stage reaction there is employed an excess of two PCR primers one of which hybridises to the 3' end of sequence x_1' and the other of which hybridises to the 3' end of the sequence x_n (and which therefore respectively provide the 5' ends of the sense and anti-sense strands of the intended hybrid molecule). The excess is a concentration of primers such that an increase in concentration does not produce an increase in yield. The "excess" may readily be determined by a person skilled in the art and for a typical reaction might for example be 500nM.

Fresh polymerising enzyme is used for the second stage PCR reaction and this is preferably a heat activated polymerase enzyme as described for the first stage reaction.

The second stage reaction may be conducted under temperature cycling conditions as described for the first stage reaction.

The invention is further described with reference to the accompanying drawings, in which:

Fig. 1 illustrates a DNA hybrid molecule and a "naturally occurring" DNA molecule from which it is produced; and

Fig. 2 illustrates steps in the conversion of the "naturally occurring" DNA molecule to the hybrid molecule.

Figs. 3 to 7 illustrate the results of the ExampleS.

The invention is described by way of example only with reference to the synthesis of a DNA hybrid molecule 1 (see Fig. 1) comprised of exons present in a naturally occurring DNA molecule 2 (see Fig. 1). More particularly, the molecule 2 is shown as being comprised of sense and anti-sense strands 3 and 4 respectively with the former incorporating exons x_1, x_2, x_3.....x_n of interest (e.g. for the purpose of mutation analysis). These sequences x_1, x_2, x_3.....x_n have their respective complementary sequences x_1', x_2', x_3'.....x_n' in the anti-sense strand 4 shown and may for example be separated from each other by several hundred bases.

The hybrid molecule 1 is illustrated as having sense and anti-sense strands 5 and 6 respectively with the former incorporating the sequences x_1, x_2, x_3.....x_n and the latter incorporating sequences x_1', x_2', x_3'.....x_n'. As shown, sequences x_1 and x_2 are connected (reading in the 5' to 3' direction) by the sequence -t-L_{12}-a- where a and t represent adenine and thymidine residues respectively and L_{12} is a linker sequence (the subscript "12" indicating that the linker is between sequences x_1 and x_2. Similarly sequences x_2 and x_3 are connected by the sequence -t-L_{23}-a and so forth.

Reference is now made to Fig. 2 which illustrates the manner in which molecule 1 is synthesised from molecule 2. More particularly, the synthesis involves a two stage PCR reaction employing heat activated *Taq* as the polymerising enzyme and in which the first stage utilises limiting concentrations of a plurality of pairs of modified PCR primers, one pair for each of sequences (x_1, x_1'), (x_2, x_2'), (x_3, x_3')..... (x_n,x_n') to be incorporated in the hybrid molecule. For each such sequence (x_j, x_j') (where j is 2 to (n-1) which is to be an internal sequence in the hybrid molecule 1 (i.e. all sequences except (x_1, x_1') and (x_n, x_n')) the primer pair comprises

(i) a first primer having a priming sequence which will specifically hybridise to the 3' region of the "sense sequence" x_j and which is connected at its 5'-end to a linker sequence $L'_{j,(j+1)}$ via an adenine residue ("a"); and

(ii) a second primer having a priming sequence which will specifically hybridise to the 3' region of the "antisense sequence" x_j' connected at its 5'-end to a linker sequence $L_{j-1,j}$ via an adenine residue ("a").

The primer pair for sequence (x_1, x_1') comprises a primer as defined under (i) above (in which the linking sequence is L_{12}) and a conventional PCR primer (P1) specific for the 3' region of sequence x_1'. The primer pair for sequences x_n, x_n' comprises a primer as defined under (ii) above (in which the linking sequence is $L_{(n-1),n}$) and a conventional PCR (P2) primer specific for the 3' region of sequence x_n.

The manner in which the described primers are intended to hybridise to the starting DNA molecule 2 (when denatured into its component strands) is as illustrated in Fig. 2.

The linker sequences which form part of the modified PCR primers are such that they do not have any internal secondary structure (e.g. "hairpins") and that two

such complementary sequences have an annealing temperature (T_m) which slightly exceeds that of the priming sequence and its complement. Examples of suitable linker sequences have been given above.

In the first stage PCR reaction, the "starting" DNA molecule 2 is treated with the sets of primers as described together with a heat activated *Taq* polymerase (e.g. having an activation temperature of ca 94°C) in an appropriate buffer. In this first stage, all of the primers are present in limiting concentration.

The first stage is conducted under conditions of temperature cycling such that there is an initial, relatively high temperature, denaturation step (e.g. at 95°C), followed by a hybridisation step (e.g. at 60°C) followed by a strand synthesis step (e.g. at 72°C).

As a result of this series of temperature cycling steps, a PCR reaction is effected resulting in the generation of the "short" products (see Fig. 2) in which the 3'-ends of all x and x' sequences (except x_1 and x_n') are connected to their respective linker sequences via an adenine residue "a" and the 5'-ends of all sequences (except x_1' and x_n) are connected to their respective linker sequences via a thymidine residue "t". Furthermore, it will be noted that the 3'-end of each of the short products has an adenine residue as added by the *Taq* polymerase.

Whilst we do not wish to be limited by any particular mechanistic interpretation, we believe that the first stage PCR reaction also results in a degree of formation of "longer" products. The manner in which such "longer" products will form from selected ones of the short products is illustrated in those steps of the reaction illustrated within the dashed box. More particularly, complementary linker sequences are able to hybridise to each other and it will be noted that their flanking "t" and "a" residues respectively hybridise to the "a" and "t" residues which flank the complementary linker sequence. Thus "self-primed" constructs are generated and

extension can occur in the direction of the arrows to produce the "longer" products. Moreover it will be appreciated that a terminal linking sequence of such longer product so generated may be able to hybridise with a terminal linking sequence of a further "longer" product or of a short product so that further extension is possible to produce fragments of greater length. (It should be noted that the generation of the longer products may take place partly in the first stage reaction and partly in the second stage reaction (discussed below) although we do not wish to preclude the possibility that these products are formed wholly in either the first or second stage reaction. It is for this reason that the generation of the "longer" products is illustrated within the box defined by dashed lines and the first and second stages of the PCR reactions are connected by dashed arrows.

At the end of the first stage reaction, the reaction mixture is preferably frozen to -20°C to deactivate any residual polymerase activity.

For the second stage of the reaction, an excess of flanking primers which define the 5' ends of the sense and anti-sense strands of the intended hybrid molecule is added to the reaction mixture. Fig. 2 illustrates these flanking primers are illustrated as FP1 and FP 2 together with the locations at which they hybridise. Also provided for the second stage reaction is further polymerase which is activated only at elevated temperature together with buffers, nucleotides etc. The reaction mixture is then subjected to temperature cycling as previously i.e. denaturation (e.g. at 95°C for 1 min), hybridisation e.g. at 60°C (e.g. for 1 min) and synthesis (e.g. at 72°C for 2 minutes).

Since the polymerase is only activated at elevated temperature, it is ensured that any sequences which randomly hybridise during the initial heating of the reaction mixture become denatured before the temperature at which the polymerase is activated so that there are substantially no extension reactions resulting from these hybridisations. Put another way, the only hybridisations which occur above the

activation temperature of the polymerase are those which are required for generation of the hybrid molecule 1 as explained more fully below.

As a result of the second stage reaction there is produced the intended hybrid molecule.

It should be appreciated that a number of modifications may be made to the illustrated protocol. For example, at least some of the sequences $x_1, x_2....x_n$ may initially be provided on separate chromosomes.

Furthermore, the primers p_1 and p_2 may be external to the sequences x'_1 and x_n respectively.

The invention is further described by the following non-limiting Example.

Example 1

1.1 **MATERIALS AND METHODS**

1.1.1 **Primer design**

A set of primers were designed to amplify exons 6, 7, 8, 9 and 10 of the Neurofibromatosis type 2 (NF2) gene (see Appendix A). These five exons comprise approximately the middle third of the coding sequence of the gene. Primers were designed with the assistance of the Oligo v 4.1 program (MedProbe, Postboks 2640, St. Hanshaugen, N-0131, Oslo, Norway). The genomic specific segments of the primers were selected to have closely equivalent T_ms (average T_m = 62.3°C @ 180mM salt concn.). The complementary 5' termini of the self assembling primers were designed with randomly selected sequence of approximately 50% GC content, with no internal secondary structure and to have T_ms that exceeded those of the genomic specific seqments in order to favour self assembly (average T_m = 64.9°C @ 180mM salt concn.). A supplementary adenine residue was inserted between the genomic and 5' complementary segments of the self assembling primers in order to

accommodate the 3' adenine overhangs added by *Taq* polymerase to the nascent DNA strand. All primers were checked for homology to *Alu* repeat sequences using the BLAST analysis program available at http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-blast. Table 1 gives the sequences and T_ms of the 12 primers used in the NF2 exon 6-10 array.

Primers were all synthesised at 0.05μM scale and were reverse phase cartridge and HPLC purified (MWG Biotech).

1.1.2 Amplification of self assembling arrays

DNA was extracted from peripheral blood lymphocytes on an Applied Biosystems 380A DNA extractor.

PCR amplification of the self assembling DNA arrays was carried out in two stages. The primary reactions were carried out in 20μl volumes using 50ng of genomic DNA, 40nmol.L^{-1} of primers 1-10 (Table 1), 750μmol.L^{-1} of each dNTP, 0.6U Platinum Taq polymerase (GibcoBRL) in a 1xPCR buffer comprising 67mM Tris-HCl (pH8.3 @ 25°C), 16.6mM ammonium sulphate, 3.7mM $MgCl_2$ and 0.085mg.mL^{-1} BSA. PCR amplification was carried out on Perkin-Elmer 2400 or 9600 thermal cyclers using the following parameters; initial denaturation 94°C (3 mins), 30 cycles of 94°C (1 min); 60°C (1 min); 72°C (2 mins) followed by a final synthesis of 72°C (10 mins). Immediately on completion of cycling the primary PCR reactions were frozen at -20°C to inactivate residual DNA polymerase activity.

Table 1: Primer sequences for amplification of NF2 exon 6-10 self-assembling array. Primers 2 through to 9 are all internal to the self assembling array and contain one of four pairs of complementary 5' termini which are highlighted in italics. The extra adenine nucleotide incorporated to accommodate the 3' terminal adenine residue added by *Taq* polymerase to the nascent strand is underlined in bold type. The melting temperature (T_m) of the respective genomic or linker portions of each of the primers is also indicated.

SEQ ID No.	Name	Length	Sequence 5'→3'	T_m genomic segment (°C)	T_m linker (°C)
01	Ex6F (ext)	23	Gtggcaaacaataccaaatttac	59.5	-
02	Ex 6R-L4F	47	*Tgtctcactgaxacctgcctacct***a**cccataaaggaatgtaaaccaac	60.7	63.7
03	Ex 7F-L4R	44	*Aggtaggcaggttcagtgagaca***a**ccgctctccacccatctcac	65.9	63.7
04	Ex 7R-L1F	42	*Agccactacccaaactcctgt***a**tggccctcactcagtctctg	61.6	61.2
05	Ex 8F-L1R	42	*Acaggagtttgggtagtggcta***g**agcctcagctggcgcttac	65.8	61.2
06	Ex 8R-L3F	43	*Tcatattagccgctgcattgcc***a**gatctgctggacccatctgc	63.5	68.0
07	Ex 9F-L3R	43	*Ggcaatgcagcggctaatatga***a**aggctgtcggactgaaactg	60.9	68.0
08	Ex 9R-L5F	49	*Cctcattaccggctgtcagactg***a**ttctcagaaaagctaccattatcag	60.4	66.7
09	Ex 10F-L5R	47	*Cagtctgacagccggtaatgagg***a**ggcagtgaagtaaatttgtggat	61.1	66.7
10	Ex 10R (ext)	20	Aggccaggactgaccacaca	65.3	-
11	Ex6F (int)	25	Catgtgtaggtttttatttgctc	60.7	-
12	Ex 10R (int)	23	Tgaccacacagtgacatcatcag	62.3	-

The secondary reactions were carried out in separate 20µl volumes using 2µL of the primary PCR, 500nmol.L^{-1} of primers 11 and 12 (Table 1), 200µmol.L^{-1} of each dNTP, 0.6U Platinum Taq polymerase (GibcoBRL) in a 1xPCR buffer comprising 50mM Tris-HCl (pH9.0 @ 25°C), 20mM ammonium sulphate and 1.5mM $MgCl_2$. The thermal cycling conditions were identical to those used for the primary amplifications.

1.1.3 Labelling of DNA

Depending upon the final application the internal primers (11 & 12) were either unlabelled (sequencing and CCM probe DNA) or 5' biotin labelled (CCM test DNA). For the production of internally labelled fluorescent CCM probe DNA, secondary PCRs were set up with primary PCRs from normal control DNA, with the addition of TAMRA labelled dCTP (Perkin-Elmer) to a final concentration of 800nmol.L^{-1} of secondary PCRs. Otherwise reaction conditions were identical to those described previously for secondary PCRs.

1.1.4 Cycle Sequencing

For sequencing the remainder of the product was then electrophoresed on a preparative 1% agarose gel and the 1046bp band was excised and purified using a QiaQuick gel purification kit (Qiagen) according to the manufacturer's protocol with the modification that the dissolved gel slice was passed through the column three times during the binding stage in order to maximise recovery efficiency. The purified sample was then eluted off the column by the addition of 40µL of 10mM Tris.HCl (pH 8.0).

40-60ng of each sample was sequenced in both orientations using either primers 11 or 12 with BigDye terminator cycle sequencing kits (Perkin-Elmer, Applied Biosystems)). The manufacturer's protocols were followed with the exception that the annealing temperature for the cycle sequencing reaction was increased from

50°C to 55°C to reflect the high T_m of primers 11 and 12. The sequencing reactions were then electrophoresed and collected on an Applied Biosystems 377 fluorescent sequencer using 48cm well to read plates.

1.1.5 Chemical Cleavage of Mismatches (CCM)

Heteroduplexes were formed by heating equal quantities of unpurified biotinylated test DNA with unpurified internally TAMRA labelled normal control probe DNA at 94°C for 5 mins followed by annealing at 65°C for 12 hours.

The heteroduplexed DNA was purified by electrophoresing each sample through a 1% low gelling temperature agarose gel and cutting out the 1046bp band using a sterile scalpel. The volume of gel slice was estimated by weighing and the gel slice equilibrated for 20 mins in an equal volume of 1x β-Agarase buffer comprising 10mM Bis-Tris HCl (pH6.5) and 1mM EDTA. The β-Agarase buffer was then removed and the gel slice heated at 70°C for 10 mins to liquefy the agarose followed by cooling to 37°C. A further equal volume of 1 x β-Agarase buffer was added followed by 1U of β-Agarase I (USB Biochemical). The gel slice was then digested overnight at 37°C.

30μL of each digested heteroduplex DNA sample was then combined with 10 μg of Dynabeads M-280 Streptavidin (Dynal), freshly washed according to the manufacturer's instructions and resuspended in 30μL of a binding and washing buffer comprising 2M NaCl, 10mM Tris.HCl (pH8.0), 0.1mM EDTA and 0.4% Tween 20. The heteroduplexes were then left to complex with the beads for 1 hour at 42°C. The samples were then placed on a magnet to separate the beads and the supernatant carefully removed using a pipette.

1.1.6 Hydroxylamine modification

The beads with bound DNA were resuspended in 20μL of a solution of 4M hydroxylamine, 2.3M diethylamine and incubated at 37°C for 2 hours. The samples were then placed on a magnet to separate the beads and the supernatant was carefully removed using a pipette. Piperidine cleavage of modified bases was then immediately carried out.

1.1.7 Potassium permanganate modification

The beads with bound DNA were resuspended in 20μL of a solution of 1mM $KMnO_4$, 3M tetraethylammonium chloride and incubated at 25°C for 10 mins. The samples were then placed on a magnet to separate the beads and the supernatant was removed using a pipette. Piperidine cleavage of modified bases was then immediately carried out.

1.1.8 Piperidine cleavage

Prior to piperidine cleavage the beads were rinsed once with 50μL of TE buffer. Then the beads were resuspended in 5μL of a 1M solution of piperidine in deionised formamide to which Genescan 2500 Rox size standard (Perkin-Elmer) and dextran blue loading dye (Sigma) had been added. The samples were heated at 90°C for 30 mins to cleave the modified bases, snap chilled on ice and placed on a magnet to separate the beads from the unbound DNA now in the liquid phase.

1.1.9 Fluorescent fragment analysis

The supernatant was loaded onto a 24cm well to read 4% denaturing polyacrylamide gel on an Applied Biosystems 373XL fluorescent analyser. The gel was electrophoresed at 2500V, 30mA and 35W for 12 hours and data was collected with filter set A. After electrophoresis each lane was tracked and size calling performed using the internal size standard as a reference. Test samples were then compared with the normal control sample for the presence of novel cleavage products by overlaying using Genescan software.

1.2 Results

1.2.1 A 2µl portion of the product of step 1.1.2 applied to a 1% agarose gel confirmed the presence of the expected 1046bp fragment.

1.2.2 PCR amplification of each individual exon from the product of step 1.1.2 under common amplification conditions gave a single fragment of the expected size, confirming the integrity of the primer syntheses and the compatibility of the primer pairs. (see Fig 3a.)

1.2.3. Figs 4a and 4b show the product of step 1.1.2 sequenced from both termini. Fig 5a shows sequencing in the forward orientation using primer 11. Fig 4b shows sequencing in the reverse orientation using primer 12. The positions of the intron exon boundaries, primer annealing sites and linker sequences are indicated on the electropherograms. The product was seen to comprise the expected exons in the correct orientations. No discernible degradation of sequence was observed across the transitions from one component to the next.

1.2.4 The procedure of step 1.1.2 was repeated for a range of genomic DNA samples and consistently yielded fragments of the expected size. Reaction yields were generally high (~500ng per 10µl) with low background (see Fig 5a).

Example 2

In order to test the suitability of the products obtained by the procedure of the invention for downstream applications like mutation scanning, products from a series of 7 heterozygotes for NF2 mutations spread throughout four of the five NF2 *CR* exons were directly sequenced to confirm that the genotype present in genomic DNA was correctly represented in the product. The 7 mutant heterozygotes were as follows

nt600-3C>G (exon 7), nt676-7T>G (exon 8), nt713delC (exon 8), nt784C>T (exon 8), nt855delT (exon 9), nt887delT (exon 10) and nt948G>T + nt949G>T (exon 10). All 7 mutations were clearly visible in the heterozygote state. Fig 6 gives example data from four representative mutation heterozygotes. We have also obtained heterozygote sequencing data of comparable quality on products that have not been gel-purified (data not shown). Thus, the gel purification step may be omitted in many protocols.
11 NF2 mutant heterozygotes were then retrospectively screened using a modified fluorescent solid-phase CCM method based on that of Rowley et al, as disclosed in Genomics 30, 574-582. The products were internally labelled using TAMRA dCTP in preference to end labelling. This method helps eliminate false positives resulting from background cleavage. Cleavage of internally labelled products should result in two labelled fragments with a total molecular weight equal to the uncleaved product. Mismatched cytosines were modified using hydroxylamine. However, on the grounds of safety and convenience mismatched thymines were modified using the potassium permanganate in preference to osmium tetroxide.

The results are shown in Fig 6 Cleavage products are clearly visible in the lanes containing products from the following mutation heterozygotes Lane 3 - nt600-3C>G (exon 7); Lane 6 - nt713delC (exon 8); Lane 8 - nt678insC (exon 8); Lane 10 - nt770delC (exon 8); Lane 11 - nt676-7T>G (exon 8) and Lane 12 - nt948G>T + nt949G>T (exon 10). The remaining samples are normal controls or from mutations that are expected to give mismatches unaffected by hydroxylamine modification.

Of the 11 mutations, 10 were detectable using either hydroxylamine or potassium permanganate modification (see Table 2).

Table 2: Detection of 11 NF2 mutation heterozygotes by fluorescent solid-phase CCM. The predicted mismatched bases/pairs of bases are given for each mutation. Positive cleavage after the two modifications is indicated as either Yes - clear cleavage, (Yes) - weak but reproducible cleavage, No - no detectable cleavage.

Mutation (Exon)	Mismatch	Hydroxylamine	$KMnO_4$
nt600-3C>G (Ex 7)	G:G + C:C	Yes	No
nt676-7T>G (Ex 8)	T:C + A:G	Yes	Yes
nt676-2A>T (Ex 8)	T:T + A:A	No	No
Nt678insC (Ex 8)	G + C	Yes	(Yes)
Nt713delC (Ex 8)	G + C	Yes	No
Nt770delC (Ex 8)	G + C	Yes	No
Nt784C>T (Ex 8)	T:G + A:C	(Yes)	No
Nt855delT (EX 9)	A + T	No	Yes
nt887delT (Ex 10)	A + T	No	Yes
Nt903C>T (Ex 10)	C:A + T:G	Yes	No
nt948G>T + nt949G>T (Ex 10)	T:C + A:G	Yes	Yes

The mutation not detected by either condition, nt676-2A>T, is predicted to produce heteroduplexes detectable only by potassium permanganate modification (T:T/A:A mismatches). Furthermore, two other mutations producing heteroduplexes with mismatched thymines (nt784C>T and nt903C>T) had undetectable levels of cleavage after potassium permanganate modification. However, both these mutations also produce mismatched cytosines which were detected after hydroxylamine modification. All of the mutations predicted to produce heteroduplexes with mismatched cytosines produced visible cleavage products after hydroxylamine modification.

Example 3

3.1 Procedure

The procedure of steps 1.1.1 and 1.1.2 (see Example 1) was repeated to amplify exons 8, 9, 10, 11 and 12 of the Human mismatch repair, hMLH1 gene. The primers are as shown in Table 3.

The genomic specific segments of the primers were selected to have closely equivalent T_ms (set average T_m = 60.8°C at a salt concentration of 180mM). The complementary 5' termini or linker segments of the primers were designed with random sequence of approximately 50% GC content, with no internal secondary structure and to have T_ms that exceeded those of the genomic specific seqments in order to favour self assembly (set average T_m = 65.4°C, both at a salt concentration of 180mM). A supplementary unmatched adenine residue was inserted between the genomic and 5' linker segments of the self assembling primers in order to accommodate the 3' adenine overhangs added by *Taq* polymerase to the nascent DNA strand. All primers were checked for homology to *Alu* repeat sequences using the BLAST analysis program available at http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-blast.

Table 3

SEQ ID No.	Name	Length	Sequence 5'→3'	T_m genomic segment (°C)	T_m linker (°C)
13	Ex 8F (ext)	21	Gatgtttcagtctcagccatg	58.7	-
14	Ex 8R-L3F	49	*Tcatattagccgctgcattgccaggaatgataaaccaagataataaatg*	59.3	68.0
15	Ex 9F-L3R	49	*Ggcaatgcagcggctaatatgaattctttttgaaatgtttgagttttgag*	60.6	68.0
16	Ex 9R-L1F	43	*Agccactacccaaactcctgtacctgtgagtggatttcccatg*	63.3	61.2
17	Ex 10F-L1R	44	*Acaggagtttgggtagtggctaccctcagacagttttgaactgg*	62.2	62.2
18	Ex 10R-L4bF	46	*Tctctcactgaatccgcctacctacttggttgaggagtttggtgct*	64.1	65.2
19	Ex 11F-L4bR	48	*Aggtaggcggattcagtgagagaaccctcccactatctaaggtaattg*	61.6	65.2
20	Ex 11R-L5bF	44	*Taacattccaggctgtcggactgaagtagctggatgagaagcgc*	60.6	67.1
21	Ex 12F-L5bR	50	*Cagtccgacagcctggaatgttaatttaatacagactttgctaccaggac*	61.7	67.1
22	Ex 12R (ext)	25	Taaagagtagctgtactttttcccaa	60.3	-
23	Ex 8F (int)	22	Taaatccttgtgtcttctgctg	58.3	-
24	Ex 12R (int)	18	Aagccatacctgggggttg	58.8	-

3.2 Results

3.2.1 A 2µl portion of the product of step 1.1.2 applied to a 1% agarose gel confirmed the presence of the expected 1247bp fragment.

3.2.2 PCR amplification of each individual exon from the product of step 1.1.2 under common amplification conditions gave a single fragment of the expected size, confirming the integrity of the primer syntheses and the compatibility of the primer pairs. (see Fig 3b.)

3.2.3 The procedure of step 1.1.2 was repeated for a range of genomic DNA samples and consistently yielded fragments of the expected size. Reaction yields were generally high (~500ng per 10µl) with low background (see Fig 5b).

APPENDIX A

Genomic sequences of NF2 exons 6-10 amplified by the self assembling array. Exonic sequence is indicated by capitals, intronic sequence is in lower case. The annealing sites of primers used in the primary PCR reaction to generate the five self assembling DNA fragments are marked in bold underlined sequence. The annealing sites of the internal primer pair used during the secondary PCR to drive the array assembly are marked in underlined italics. The primary and secondary primers for exon 10 reverse overlap, the region of overlap is marked in bold, underlined italics.

NF2 Exon 6 (SEQ ID NO. 25)

tctgtgtgac tactcctggt gtagctttaa aatagcttta ctgtttgtaa aatgatgcat 60
aattatagaa gtggcaaaca ataccaaatt tacttcatgt gtagattttt tattttgctc 120
tatttttggg tagGTAATAA ATCTGTATCA GATGACTCCG GAAATGTGGG AGGAGAGAAT 180
TACTGCTTGG TACGCAGAGC ACCGAGGCCG AGCCAGgtga ggcccattca ttgttggttt 240
acattccttt atgggc 256

NF2 Exon 7 (SEQ ID NO. 26)

gaatgcttga tttggtggcc caccagctct ccacccatct cacttagctc caatgacagt 60
gtcttccgtt ctccccacag GGATGAAGCT GAAATGGAAT ATCTGAAGAT AGCTCAGGAC 120
CTGGAGATGT ACGGTGTGAA CTACTTTGCA ATCCGGgtgt gttgaaacct ctctgagctc 180
cttgtgtagt agacagagac tgagtcaggg ccaggactgc taaaatggtt acttcttcat 240

NF2 Exon 8 (SEQ ID NO. 27)

tctgtggacc tgctgaactg cacatgtgac agtgtgtgcc agattctttg gaaggttgaa 60
taaaatttt gagcctcagct ggcgcttaca gtagctgttc ttattggatc cacagAATAA 120
AAAGGGCACA GAGCTGCTGC TTGGAGTGGA TGCCCTGGGG CTTCACATTT ATGACCCTGA 180
GAACAGACTG ACCCCCAAGA TCTCCTTCCC GTGGAATGAA ATCCGAAACA TCTCGTACAG 240
TGACAAGGAG gtaggacatg tgtgtactgc agatgggtcc agcagatctt tccctgtctg 300
ccccccttac tggagcctcc ccagccaggg catctccttg ttattcatag agtcctttaa 360
ttcccaggct ttgagggtgt ggttgtt 387

NF2 Exon 9 (SEQ ID NO. 28)

gacttggtgc tcctaattcc ctgaggttta gtgcctggat actgggaagc cagnacaagg 60

gcataacntc atgctggtct gtggccagtg tggttgcgca tttgtggaat tnccaattgc 120

tggtaacatt ccaggctgtc ggactgaaac tgtgttctgc ttcattcttc cagTTTACTA 180

TTAAACCACT GGATAAGAAA ATTGATGTCT TCAAGTTTAA CTCCTCAAAG CTTCGTGTTA 240

ATAAGCTGGT AAGTTGAGAT CCTGgtaagt tgagatcctg gttttcatta ctgataatgg 300

tagcttttct gagaa 315

NF2 Exon 10 (SEQ ID No. 29)

tgctacctgc aagagctcaa actgctatgg cactagtggg ccagta ggcagtgaagtaaa 60

tttgtgggta ttaacctttt tgtctgcttc tgtggccaca gATTCTCCAG CTATGTATCG 120

GGAACCATGA TCTATTTATG AGGAGAAGGA AAGCCGATTC TTTGGAAGTT CAGCAGATGA 180

AAGCCCAGGC CAGGGAGGAG AAGGCTAGAA AGCAGgtgag cacaaccttg ttttaactaa 240

tgatgtcact gtgtggtcag tcctggcct 269